Journal of Engineering Science and Technology
Special Issue on SOMCHE 2014 & RSCE 2014 Conference, January (2015) 86 - 95

BINDING MODE STUDY OF CELLOPENTAOSE IN β-GLUCOSIDASE B VIA DOCKING SIMULATION

N. B. AHMAD KHAIRUDIN*, N. S. F. MAZLAN

Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia
Kuala Lumpur, Jalan Semarak, 54100, Wilayah Persekutuan, Malaysia
*Corresponding Author: r-bahiah@utm.my

Abstract

Paenibacillus polymxca β-glucosidase B (BglB), belong to a GH family 1, is a monomeric enzyme that acts as an exo-β-glucosidase hydrolyzing cellobiose and cellooligosachharidess of higher degree of polymerization by cleaving β-1,4 glycosidic linkage exist between glucosyl residue. This study reports on binding modes between BglB with cellopentaose which consist of five glucosyl residue. Several visual inspection and protein –ligand interaction analysis was focused on finding amino acid residue involve in each glucosyl residue.Computational docking calculation was performed using program GOLD generating ten solution BglB-cellopentaose complexes. From visual inspection, subsite -1 record the most interacting residue namelyGln22, Asn166, Glu167, Asn296, Glu356, Trp402, Glu409 and Trp410. The reducing glycosyl at subsite +1 are making interacting with residue namely Trp328, Asn223 and His181. Meanwhile, the residues Arg243, Leu174, Gln316 and Tyr169 making contact with reducing glycosyl at the subsite +2 and subsite +3. Lastly, only two residues, His318 and Glu180 reported to make interaction within reducing glycosyl at subsite +4 as it is already over exposed towards outside of binding cleft. From overall, a total of 11 hydrogen bonds were observed in BglB-cellopentaose complex

Keywords: Cellopentaose,β-glucosidase, Binding mode, Docking.

1. Introduction

β-glucosidase (EC 3.2.1.21) from family Glycosidase Hydrolase 1 (GH1) is heteregenouse group of enzymes which are capable of hydrolyzing the β-glycosidic bonds of disaccharides, oligosaccharides, or conjugated glucosidases [1]. It commonly found in plants, fungi, mammals and microorganisms in which they play a variety of

Nomenclatures

$S_{hb\ ext}$	Protein-ligand hydrogen bond score
$S_{hb\ int}$	The contribution to the fitness score
$S_{vdw\ ext}$	The protein-ligand van der Waals score
$S_{vdw\ int}$	The contribution due to intermolecular strain in the ligand

Greek Symbols

Å	Angstrom
β	Beta

Abbreviations

3D	Three dimensional structures
Arg	Arginine
Asn	Asparagine
BglB	β-glucosidase B
CID	Chemical Identification
GH1	Glycosidase Hydrolase 1
Gln	Glutamine
Glu	Glutamic Acid
GOLD	Genetic Optimization for Ligand Docking
His	Histidine
Leu	Leucine
PDB	Protein Data Bank
PLIF	Protein Ligand Interaction Fingerprint
Tyr	Tyrosine

functions. All Family GH1 of β-glucosidase shares similar hydrolyticmechanism cleaving β-glycosidic linkage between an anomeric carbon and glycosidic oxygen. There are two active site amino acid residues, usually glutamates, play key roles in hydrolytic mechanism, with one acting as a nucleophilic and other as a proton donor [2]. These two catalytic carboxylates are highly conserved in family GH1 and also for other GH families. These enzymes are thought to work through a double displacement mechanism [3], whereby they had two step, glycosylation and deglycosylation. Both steps involved the formation of an enzyme-substrate complex. In the first step, the catalytic nucleophile attacks the anomeric carbon of the glucosyl unit to form an enzyme-glucosyl covalent intermediate while the glycosidic oxygen is protonated by the catalytic acid-base. In second step, the same catalytic acid-base receives a proton from water molecule and then attacking the anomeric carbon. Finally, the glucose monomer is released from enzyme-glucosyl intermediate.

There are many experimental studies on the catalytic mechanism used by these enzymes and also a lot of crystallographic structure have been obtained [4]. β-glucosidase B (BglB) from *Paenibacillus polymxca* [5] has been chosen as a model for this simulation because it is small (454 amino acids) and no metal ions are required for maintaining the structure stability. BglB is a retaining glycosidase enzyme that specifically binds to β-1,4 linked glucose residues in the binding site. The catalytic amino acids are Glu356 (nucleophile) and Glu167 (acid/base catalyst)[5]. These glutamate residue is important for reducing the energy barrier

of the glycosylation step [6]. It is present in BglB active pocket site and it is believed to be actively involved in substrate binding process.

Cellopentaose was chosen as substrate to investigate the key residue involve in five glucosyl residues unit as illustrated in Fig. 1. The β-glucosidase active site can be divided into several subsites with sufficient space to bind a monossacharides unit. The region of the enzyme that binds one residue of the chain is called subsite. The subsite that binds the monosaccharides of the substrate nonreducing end is called subsite -1 (glycone subsite) whereas the remaining part of the substrates is accommodated in the aglyclone binding region which may be formed by several subsites +1, +2, +3, and so on. The substrate cleave point is between subsite -1 (glycone subsite) and +1 (aglycone subsite) which is also known as glycosidic bond. Subsite -1 plays important role as it holds the sugar ring that bears the scissile bond. It also contains anomeric carbon atom which is the center of the hydrolysis reaction.

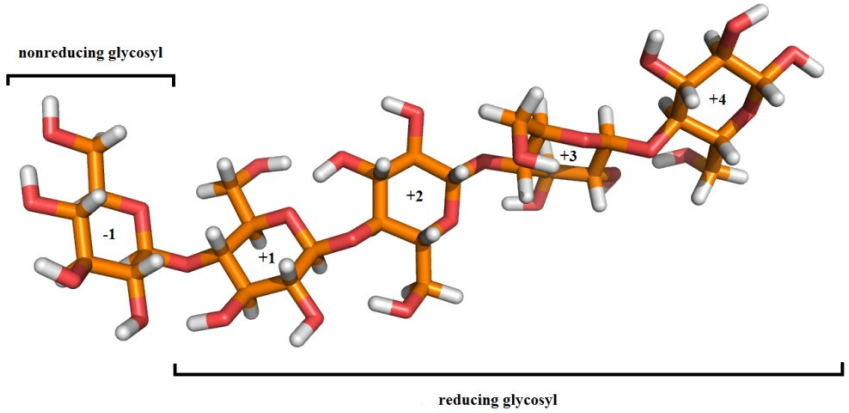

Fig. 1. Streoview of Cellopentaose Labeled with Subsite.

The molecular docking simulation by GOLD program [7] had been performed to study the binding mode of cellopentaose in BglB protein. A large number of conformations of protein ligand complexes can be generated by this program. Each conformation is positioned at active site in a variety of orientation known as pose whereas poses were ranked by the scoring function to determine the best overall poses. The scoring function incorporated in GOLD uses classical molecular mechanic pose force field that made up with four terms:(i) Protein ligand hydrogen bond energy (external H-bond), (ii) Protein-ligand van der Waals energy (external vdw), (iii) Ligand internal van der Waals energy (internal vdw) and (iv) Ligand intramolecular hydrogen bond energy (internal H-Bond). By default, the external vdw score is multiplied by a factor of 1.375 during the computation of total fitness score. Equation (1) shows the fitness function that has been optimized for the prediction of ligand binding position.

$$\text{Gold Score Fitness} = S_{hb\_ext} + 1.375S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} \quad (1)$$

where S_{hb_ext} is the protein-ligand hydrogen bond and S_{vdw_ext} is the protein-ligand van der Waals score . S_{hb_int} is the contribution to the fitness due to the intermolecular hydrogen bonds. S_{vdw_int} is the contribution due to intermolecular strain in the ligand [8] .

2. Material and Methodology

2.1. Material

2.1.1. Protein and ligand structure

The crystallographic structure of β-glucosidase B was obtained from PDB website (http://ww.pdb.org) PDB's ID 2O9T [5] having resolution of 2.15Å . The two dimensional (2D) ligand structure of cellopentaose was obtained from PubChem (http://www.pubchem.org) with CID 440949. It consists of five glucosyl residue with chemical formula $C_{30}H_{52}O_{26}$.

2.1.2. Softwares

A LowModeMD module from program MOE 2011 version 10 [9] was used to built 3D structures and optimize ligands structure prior to docking simulations. It also used to visualize protein-ligand complexes structure and map out hydrophobic interactions.GOLD version 5.1 [8] was used for performing docking calculation. The graphical display results were generated using Pymol version 1.3[10] from Schrodinger.

2.1.3. Hardwares

The ligands optimization were carried out in Macintosh machine running on Intel Core i5 processor with 2GB RAM and 500GB hard disk with Mac OSX Lion version 10.7.5 as the operating system. The computational docking was performed in machine running on Intel Core i7 processor with 8GB RAM and 500GB hard disk with CentOS version 5 as the operating system.

2.2. Methodology

2.2.1. Preparation of protein and ligand structure

The originally bound ligand, water and cofactors in BglB structure were being removed from the protein file. The new BglB structure was saved in *.pdb* format to be use later in docking simulation. The three-dimensional (3D) structure for cellopentaose was built and optimized using the Builder module in MOE 2011.10. The multi conformation structural data of the ligands was generated using LowMode MD in MOE 2011.10 using default parameters setting: The programcan generate hundreds of possible conformations of 3D structure. From the results, only the top ten of ligands conformations were selected and saved in *.mol2* format to be used in docking simulation.

2.2.2. Protocol of docking simulation

The BglB and cellopentaose structures were loaded into the program GOLD. The binding site was set manually at coordinate 66, 28, and 38 for x, y, and z respectively. The active site was defined within 10 Å all heavy atoms of ligands in the complexes and ligand-binding interactions were analyzed using Gold Score (GS) scoring function. The default settings were applied for all parameters stated: population size (100); selection-pressure (1:1); number of operation (10,000); number of island (1); niche size (2) and operator weight for migrate (0), mutate

(100) and crossover (100) .The simulation generated 100 of conformations pose of protein-ligand complex that were ranked according to GOLD Score fitness. The option for early termination of the simulation was turn on if very similar results were obtained. In the end, the top ten conformations pose for each substrate were saved in *.pdb* for further analyses.

2.2.3. Analysis of docking result

All the ten BglB-cellopentaose complexes were subjected for visual inspection using PyMol and MOE programs. The Protein Ligand Interaction Fingerprint (PLIF) and Ligand Interaction modules in MOE were used to identify the important acid amino residue that involved in the binding mode of cellopentaose in BglB.

3. Results and Discussion

3.1. GOLD fitness scores

The full information of list of fitness score for the top ten BglB-cellopentaose complexes were listed in Table 1.The average for fitness score is 50.571,$S_{hb_ext}$is 19.594, $S_{vdw_ext}$is 23.51, and $S_{vdw_int}$is -1.351. The highest fitness score was recorded at 56.55. All the result obtained look satisfying since there is no much difference between the score for each docked complexes. The view of the superimposed of all the ten docked pose conformation was illustrated in Fig. 2. The subsite -1 was observed buried into the deep cleft of the BglB enzyme. The subsite +3 and +4 were seemed to be overexposed to outside of the binding pocket since the cellopentaose has a long chain. In general, the overall structure of cellopentaose has twisted linear structure due to the strong hydrogen bond formed between the neighboring glucosyl residues [11].

Table 1. List of GOLD Score of Substrates Binding with BglB Enzyme.

Complex	Fitness Score	S_{hb_ext}	S_{vdw_ext}	S_{vdw_int}
1	56.55	21.84	25.81	-0.78
2	56.24	21.93	25.44	-0.68
3	54.17	21.77	24.76	-1.64
4	54.14	21.83	25.11	-2.21
5	54.04	21.88	24.99	-2.21
6	47.68	17.67	22.37	-0.75
7	47.12	18.47	21.81	-1.35
8	46.42	17.61	21.45	-0.68
9	45.23	17.22	21.26	-1.23
10	44.12	15.72	22.10	-1.98

3.2. Molecular data interactions

All the ten complexes structures from docking simulation were imported into the software MOE Tools. The software calculated the molecular interaction and generates two-dimensional (2D) pictures that gave a better view for quick inspection of each complex. The results indicated there are large number of hydrogen bond interaction established between cellopentaose and amino acid

residue in the binding sites of BglB. A total of 11 hydrogen bonds had been recorded as listed in Table 2.

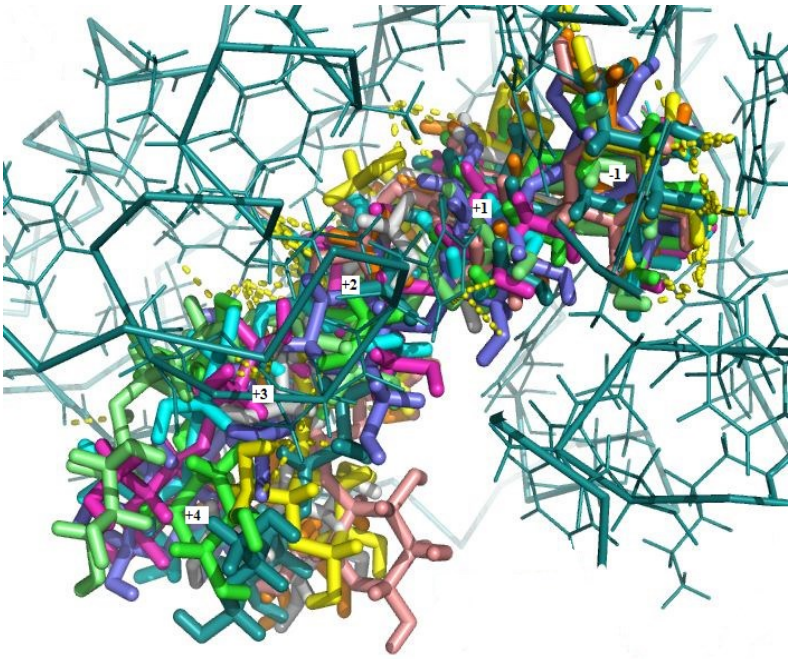

Fig. 2. Superimpose of the ten BglB-Cellopentaose Complexes Conformations Generated from Docking Simulation.The Hydrophobic Interaction between Enzyme and Glucosyl Residue are Indicated by Yellow Dashed Line.

Table 2. List of Hydrophobic Interaction between BglB and Cellopentaose.

Ligand Atom	Receptor Atom	Distance (Å)	Energy (kcal/mol)
O4	SG CYS170	4.10	-0.6
O9	OE1 GLU225	2.40	2.1
O12	OE2 GLU167	2.47	-1.6
O20	OE2 GLU409	2.55	-3.2
O26	OD1 ASN166	2.62	-2.3
C38	OE1 GLN316	3.16	-0.8
C56	OE2 GLU356	2.97	-1.3
O11	NH1 ARG243	2.73	-2.5
O19	NE2 GLN316	3.25	-0.8
O20	NE2 GLN22	3.38	-0.7
O22	NE1 TRP410	2.84	-2.9

From the interaction data obtained, it clearly demonstrate that the binding of cellopentaose depend on the hydrophobic interaction involved between the enzymes residue and the substrates. Some of the interaction are well conserved among retaining glycosidase family especially the hydrogen bridge between the nucleophile and the C_2-OH group of the glucose ring at the subsite -1 [12]. As

shown in Fig. 3, the nucleophilic Glu356 making interaction with C56 (C_6OH subsite -1) and the proton donor interact with O12 (C_3OH subsite +1). Other glutamate residue, Glu409 formed strong hydrogen bond with O20 (C_3OH subsite -1).

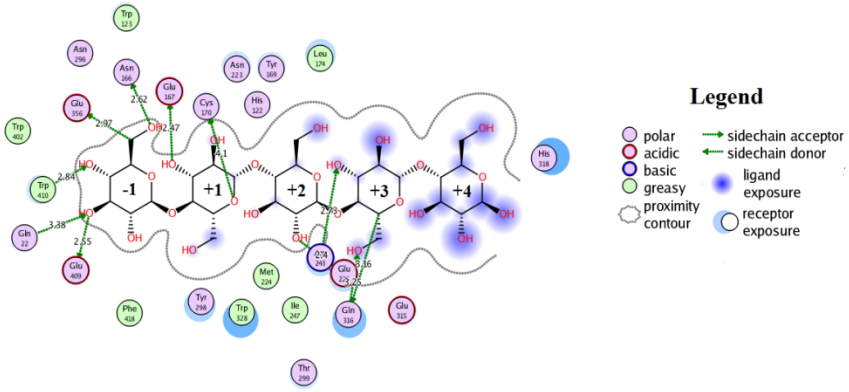

Fig. 3. The 2D Ligand Interaction Diagrams of BglB-Cellopentaose.

Several other preserves aromatic residues such as Trp, Phe and Tyr were observed stay enclosed within the subsite -1which are common in carbohydrate-protein interaction [13].These conserve amino acids shows similar function and display a position and orientation, which promotes the packing of the sugar rings during the catalytic mechanism [14]. Trp410 formed hydrogen bond with O22 (C_4OH subsite -1).There are not much interaction recorded at subsite +2 and subsite +3. Arg243 was recorded making interaction with O11 (C_2OH subsite +2), Gln316 interacts with O19 (C_6OH subsite +3) and C38 (C_5 subsite +3). There is no molecular interaction recorded for subsite +4 but there are two residues are stay enclosed which are Glu335 and His338. The stereo view displays for the best BglB-cellopentaose docked pose was illustrated in Fig. 4.

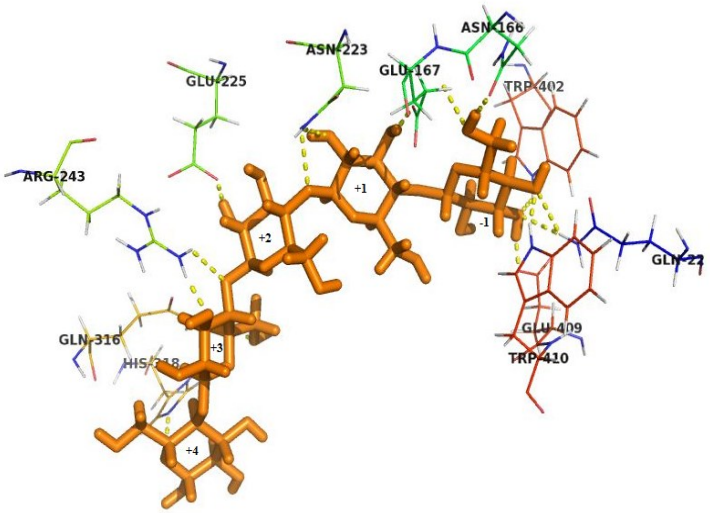

Fig. 4. Stereo View of the Best Bglb-Cellopentaose Docked Conformation where the Hydrogen Bonds were Indicated by Yellow Dashed Lines.

3.3. Protein ligand interaction fingerprint analysis

By gathering the molecular information of all the ten complexes, the common acid amino residues of the enzyme involved in the binding sites can be recognizes.The PLIF program was used to determine which amino acids residues in the enzyme active site were important during the molecular recognition process. All the residues that make direct contact with any atom of the substrates were considered as the active residue. Fig.5 shows a 2D image of cellopentaose labelled with different colour for different residue for all ten complexes.

From these data, the active residue involve for specific glucosyl residues can be summarized in Table 3. As expected, the subsite -1 had the most interacting residues follow by subsite +1, subsite +2,subsite +3 and subsite+4. Apparently, the role of important residues in catalytic mechanisms such as Glu167, Trp410, Glu256 and Tyr298 are revealed in previous crystallographic study. Glu167 is known as as a protonated agent of interglycosidic oxygen atom in hydrolysis mechanism and nucleophilic Glu356 act as stabilizer agent in the transition state [5]. Whilst, Tyr298 which is hydrogen bonded to Glu356 involved in recognition of the inhibitor[5]. In significant with the previous study, the residues namely Glu225, Tyr169, Arg243 and Gln316 shows great variability at the end of subsite +2 and subsite +3. Finally, it is remarkable that Glu180 and His318 have hydrogen bond within at subsite +4 with O3 and O5 respectively.

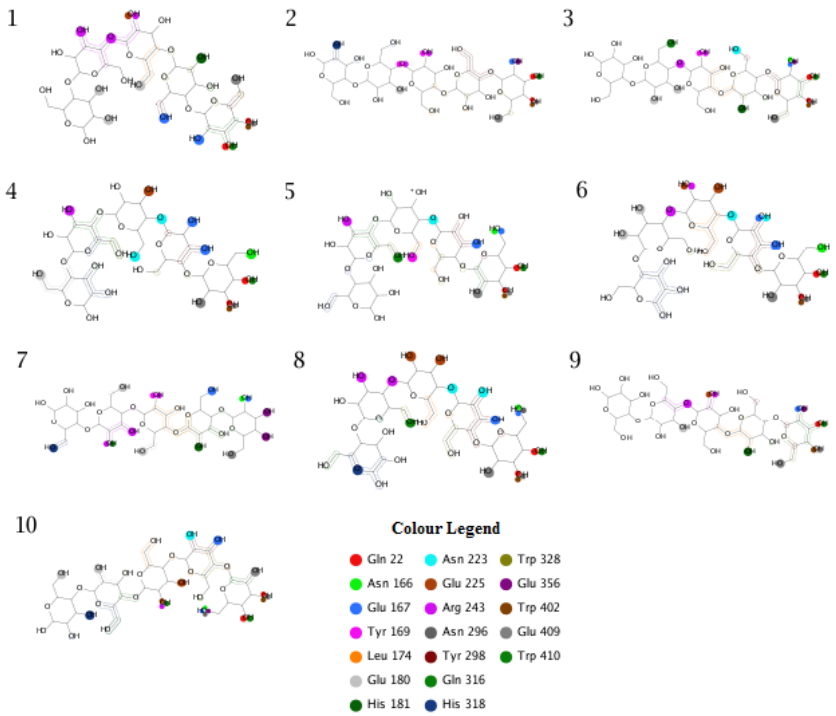

Fig. 5. The PLIF Diagram from Ligand View.

Table 3. List of PLIF According to Subsite.

Subsite	Protein Ligand Interaction Fingerprint (PLIF))
-1	Glu167, Gln22, Trp410, Glu409, Trp402, Tyr298, Asn296, Asn166
+1	Glu167, His181, Asn296, Asn166
+2	Glu225, Tyr169, Arg243, Gln316
+3	Tyr169, Glu180
+4	Glu180, His318

4. Conclusions

The molecular docking studies were conducted in order to further understanding the binding modes of cellopentaose in BglB. The docking analysis resulted in identification of important residue in protein-ligand interaction for each glucosyl residues. From the visual inspection and PLIF data, subsite -1 record the most interacting residue namelyGlu167, Gln22, Trp410, Glu409, Trp402, Tyr298, Asn296, and Asn166. The reducing glycosyl at subsite +1 are making interacting with residue namely Glu167, His181, Asn296, and Asn166. Meanwhile, the residues Glu225, Tyr169, Arg243, Gln316 Tyr169, and Glu180 making contact with reducing glycosyl at the subsite +2 and subsite +3. Lastly, only two residues, His318 and Glu180 reported to make interaction within reducing glycosyl at subsite +4 as it is already over exposed towards outside of binding cleft. Hopefully this result gave a significant insight into the catalytic mechanism of BglB towards cellooligosachharides that might be useful in designing a better β-glucosidases enzyme.

Acknowledgement

The author would like to thank Prof. Shunsuke Aoki from Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Fukuoka, Japan for guidance and providing facilities in his laboratory. The author would also like to thank MJIIT for providing the scholarship.

References

1. Esen, A.; and Blanchard, D.J. (2000). A specific beta-glucosidase-aggregating factor is responsible for the beta-glucosidase null phenotype in maize. *Plant Physiology*, 122(2), 563-572.
2. Vuong, T.V.; and Wilson, D.B. (2010). Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnology and Bioengineering*, 107(2), 195-205.
3. Henrissat, B.; and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, (318), 695-696.
4. Withers, S. (2001). Mechanisms of glycosyl transferases and hydrolases. *Carbohydrate Polymers*, 44(4), 325-337.
5. Isorna, P.; Polaina, J.; Latorre-García, L.; Cañada, F. J.; González, B.; and Sanz-Aparicio, J. (2007). Crystal structures of Paenibacillus polymyxa beta-glucosidase B complexes reveal the molecular basis of substrate specificity

and give new insights into the catalytic machinery of family I glycosidases. *Journal of Molecular Biology*, 371(5), 1204-1218.

6. Tiwari, M.; Lee, K.-M.; Kalyani, D.; Singn, R.; Kim, H.; Lee, J.-K.; and Ramachandran, P. (2012). Role of Glu445 in the Substrate Binding of β-Glucosidase. *Process Biochemistry*, 47(12), 2365-2372.
7. Luo, W.; Pei, J.; and Zhu, Y. (2010). A fast protein-ligand docking algorithm based on hydrogen bond matching and surface shape complementarity. *Journal of Molecular Modeling,* 16(5), 903-913.
8. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; and Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins*, 52(4), 609-623.
9. Molecular Operating Environment (MOE), 2013.08. (2013).
10. Schrondinger, L. (2010). The PyMOL molecular graphics system, Version 1.3.
11. Sansenya, S.; Maneesan, J., and Cairns, J.R.K. (2012). Exchanging a single amino acid residue generates or weakens a +2 cellooligosaccharide binding subsite in rice β-glucosidases. *Carbohydrate Research*, 351, 130-133.
12. Ketudat Cairns, J.R.; and Esen, A. (2010). β-Glucosidases. Cellular and molecular life sciences. *Cellular and Molecular Life Sciences*, 67(20), 3389-4005.
13. Qasba, P.K. (2000). Involvement of sugars in protein–protein interactions. *Carbohydrate Polymers*, 41(3), 293-309.
14. Chuenchor, W.; Pengthaisong, S.; Robinson, R.C.; Yuvaniyama, J.; Svasti, J.; and Cairns, J.R.K. (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *Journal of Structural Biology*, 173(1), 169-179.